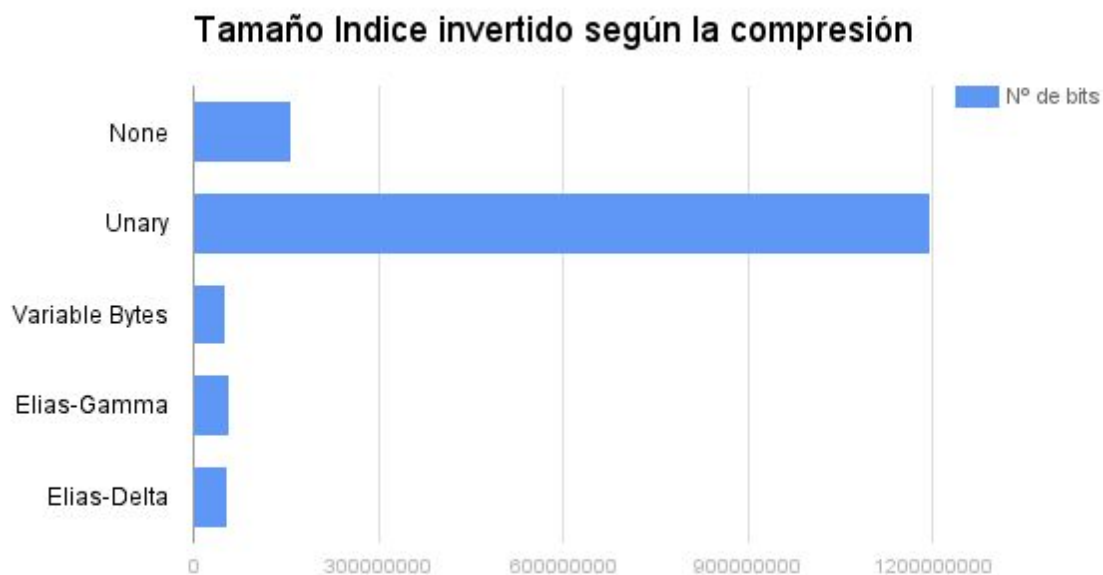


Este documento ha sido generado para la asignatura SGD  
Practica 2 Minería de datos y recuperación de la información, Ejercicio 4  
Autores: **Antonio Calvo Morata y Carlos Congosto Sandoval**

*Antonio Calvo Morata y Carlos Congosto Sandoval declaramos que esta solución es fruto exclusivamente de nuestro trabajo personal. No hemos sido ayudados por ninguna otra persona ni hemos obtenido la solución de fuentes externas, y tampoco hemos compartido nuestra solución con nadie. Declaramos además que no hemos realizado de manera deshonesto ninguna otra actividad que pueda mejorar nuestros resultados ni perjudicar los resultados de los demás.*

Los resultados obtenidos en cuanto a los tamaños del índice invertido según el algoritmo de compresión han sido:

None	158130336
Unary	1195736324
Variable-bytes	52706696
Elias-gamma	59811281
Elias-delta	54479235



Con la colección dada para este ejercicio el mejor resultado lo obtenemos con el algoritmo de bytes variables. Por lo que en este caso es el mejor algoritmo que podemos utilizar a la hora de reducir el tamaño del índice invertido generado. Debido a que hay una gran cantidad de documentos y no conocemos la estructura de éstos o que tengan un máximo de palabras y puede haber ficheros demasiado grandes (de decenas de miles de palabras), haciendo que los enteros que representan la diferencia de posiciones sean muy grandes. Esto hace que el algoritmo que mejor funcione, debido también a que no conocemos la colección, sea el de bytes variables ya que en números muy altos el tamaño que ocupa cada uno varía muy poco, podríamos decir que tiene una cota bien definida en la que los números no van a ocupar más de cierta cantidad y es más baja que la de los demás algoritmos.

Esto se ve en la gráfica mostrada en clase:

